

ABSTRACT OF THE DISCLOSURE

A computer implemented method of searching genetic data or information for a plurality of query sequences in a set of target sequence fragments, allowing for mismatches at up to n sequence positions, including dividing each query sequence of the plurality of query sequences into $n+1$ query sequence segments and dividing each target fragment of the target sequence fragments into at least $n+1$ target sequence fragment segments, for each query sequence, constructing a first query group and a second query group by distributing query sequence segments there between such that at least n query sequence segments are contained in the second query group, constructing from each target fragment a first target group having a same distribution as the first query group, and for each query sequence, comparing the first query group with each first target group to identify potential matching target fragments.